Communications in Applied Geometry.
ISSN 2249-4286 Volume 14, Number 1 (2024) pp. 5-8

www.ripublication.com/cag.htm

The Correlation Coefficient: A Geometric View from N-Space

David A. Swanson

University of California Riverside, Riverside CA 92521 and
Center for Studies in Demography and Ecology
University of Washington
Seattle, Washington USA 98195

Abstract

A number of ways to view the correlation coefficient has been identified. This paper provides a new perspective, one based on Euclidian Geometry. This perspective also can be used in statistical education, providing students with a hitherto overlooked geometric approach to undertanding correlation and regression.

Keywords: random variable, n-space, symmetry, asymmetric

Launching from a paper by Rodgers and Nicewander (1988), Rovine and Van Eye (1997) presented a 14^{th} way to look at the correlation coefficient, $p(X_1, X_2)$, where X_1 and X_2 are random variables and $p(X_1, X_2) = r$, where $-1 \leq r \leq 1$. Using a geometric interpretation of the correlation coefficient (Swanson,1977), I present yet another way to look at the correlation coefficient - by viewing it from the perspective of Euclidian N-Space.

Following Swanson (1977), let the sum of the N values of random variable X equal Σn_i. These same N values form the N elements of vector V_1. There are exactly N -1 additional vectors, V_2, V_3, …,V_s, that can be formed by permuting the N elements of V_1 such that $V_1 + V_2 + V_3 +\ldots+ V_N = (\Sigma ni_1, \Sigma ni_2, \Sigma ni_3, ,\ldots, \Sigma ni_s) = V_s$. Multiplying V_s by the scalar (1/N) gives V_m, the point in N-space that is $(1/N)^{th}$ the distance from the origin to V_s. Let D be the Euclidian distance between V_1 and V_m. By multiplying D by the scalar $(1/N)^{½}$, the standard deviation of random variable X can be presented as $\sigma = (1/N)^{½}D$.[1]

Now, let $D_{12} = ((D_1 + D_2) + (\varepsilon D_1 D_2))^{½}$

where

$$\varepsilon = 2r$$

$$D_i = (1/N)^{-½}\, \sigma_i$$

$$D_{12}^2 = (1/N)\, \sigma_{1+2}^2$$

and

$$p(X_1, X_2) = D_{12} = ((D_1 + D_2) + (\varepsilon D_1 D_2))^{½}.$$

This N-space representation can be used to show the following relationships

(1) If $p(X_1, X_2) = +1$

then $D_{12} = ((D_1 + D_2) + (\varepsilon D_1 D_2))^{½}$

(2) If $p(X_1, X_2) = 0$

then $D_{12} = ((D_1 + D_2)^{½}$

(3) If $p(X_1, X_2) = -1$

then $D_{12} = ((D_1 + D_2) + (-\varepsilon D_1 D_2))^{½}$

This interpretation adds to the contribution by Falk and Well (1997) aimed at enhancing statistical education in that they give no mention of a geometric interpretation of the correlation coefficient.

It is worthwhile to note that Dodge and Roussan (2021) show that the cube of a correlation is not symmetric in X and Y (where X = X_1 in the notation given above and Y = X_2) under certain assumptions, which leads them to suggesting that one could select the associated regression model for which the response variable has the smallest skewness (i.e. select $Y_i = a_{YX} + b_{YX}X_i + \varepsilon_i$ as the model if Y has a smaller skewness than X, and select $X_i = a_{XY} + b_{XY} Y_i + \varepsilon'_i$ as the model if X has a smaller skewness than Y)

Endnote

1. Because $D = [\Sigma\ (pi - qi)2]^{½}$, where p and q are the Cartesian coordinates in N-dimensional Euclidian space and i = 1 to N, it requires no formal proof that $D(1/N)^{½}$ can be interpreted as the standard deviation of V_m (see, e.g., Spiegel,1961: 70; Swanson,2023).

References

[1] Dodge, Y., and V. Roussan (2021). From Correlation to Direction Dependence Analysis, 1888-2018. pp. 3- 7 in W. Wiedermann, D. Kim, E. Sungur, and A. van Eye (eds). *Directional Dependence in Statistical Modeling: Methods of Analaysis*. Wiley. Hoboken, NJ.

[2] Falk, R., & Well, A. (1997). Many faces of the correlation coefficient. *Journal of Statistics Education* 5(3): 1–18.

[3] Rodgers. J. & W. Nicewander (1988). Thirteen Ways to Look at the Correlation Coefficient, *The American Statistician*, 42:1, 59-66.

[4] Rovine, M. & A. von Eye (1997). A 14^{th} Way to Look at a Correlation Coefficient: Correlation as the Proportion of Matches, *The American Statistician*, 51:1, 42-46.

[5] Spiegel, M. (1961). Theory and Problems of Statistics. Schaum's Outline Series. McGraw-Hill. New York.

[6] Swanson, D. (1977). An Alternative Geometric Representation of the Correlation Coefficient.*The American Statistician* 31 (February): 52.

[7] Swanson, D. (2023). A Statistical Margin of Error from a Geometric Perspective. *Communications in Applied Geometry* 13(2): 185-186.